Genomics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Methods

Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression

Hang He [a,b,1], Huiyong Zhang [b,1], Xiangfeng Wang [c], Nicholas Wu [d], Xiaozeng Yang [d], Runsheng Chen [a], Yi Li [b], Xing Wang Deng [c,*], Lei Li [d,*]

[a] Bioinformatics Laboratory, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[b] Peking-Yale Joint Center of Plant Molecular Genetics and Agrobiotechnology, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, China
[c] Department of MCD Biology, Yale University, New Haven, CT 06520, USA
[d] Department of Biology, University of Virginia, Charlottesville, VA 22904, USA

ARTICLE INFO

Article history:
Received 26 May 2010
Accepted 28 July 2010
Available online xxxx

Keywords:
Gene expression
Tiling microarray
Allele
Hybrid
Rice

ABSTRACT

In the study of gene expression, it is often desirable to distinguish transcript pools derived from different alleles present in the same organism. We report here an oligonucleotide tiling microarray designed to specifically target 518 single nucleotide polymorphisms (SNPs) between the two sequenced rice (*Oryza sativa*) subspecies *indica* and *japonica*. The tiling array included all 25-mer probes interrogating each SNP by placing the polymorphic site at all 25 possible positions within the probe. Through hybridization to a titration series in which the *japonica*- and *indica*-derived cDNA templates were mixed with altering proportions, a regression model was used to screen for diagnostic probe sets for each SNP. Our result indicates that 284 (55%) SNPs have at least one diagnostic probe pair suitable for distinguishing and quantifying the relative abundance of allele-specific transcripts. As a proof-of-concept, we analyzed allele-specific expression in reciprocal *indica*×*japonica* F_1 hybrids and detected imbalanced expression at approximately one third of the SNPs. These results were validated by RNA-sequencing and allele-specific real-time PCR experiments. Together, our work demonstrates the utility and advantages of the tiling array method in interrogating large numbers of SNPs for quantifying allele-specific gene expression.

1. Introduction

Elucidation of the changes in gene expression associated with biological processes and developmental programs and understanding the underlying mechanisms have been a central theme in biology. Advances in molecular and computational biology in recent years have led to the development or improvement of methods for analyzing global gene expression with ever increasing experimental throughput [1,2], transcriptome coverage [3–5], and cellular resolution [6,7]. In most of these efforts, it is assumed that alleles of different origins contribute equally to the transcript pool and hence only the sum was measured. However, without allele-specific information, interpretation of gene expression could be complicated by allele-specific variation (*cis* effect), variation in other regulatory genes (*trans* effect), as well as environmental influences [8–10]. Results generated under this experimental setting thus do not offer many mechanistic cues regarding the complex allelic interactions that occur when more than one set of alleles are present in the same cell.

In diploid eukaryotic organisms, it is clear that many genes are not equally expressed from the paternal and maternal chromosomes. At the extreme are the imprinted genes that are exclusively transcribed from the non-silenced parental chromosome. We use the term "imbalanced allelic expression" (IAE) hereafter to describe variation in gene expression where alleles of the same gene are not expressed equally at the mRNA level. IAE appears to be common in heterozygous individuals [11]. For example, studies in human revealed that up to 50% of the investigated genes may exhibit IAE in heterozygote [12–14]. In plants, investigations of the transcript levels of small sets of genes indicate that IAE is potentially prevalent in heterotic F_1 hybrids [10,15,16]. These results provide a useful readout for pinpointing regulatory polymorphisms residing on the same DNA molecule that are important for controlling proper gene expression.

In angiosperm plants, polyploidy has been a prominent force in the evolution of genome organization and gene expression regulation [17]. In particular, recently formed allopolyploids typically retain duplicated copies of most genes on homeologous chromosomes that share a very high degree of sequence similarity. Numerous studies in polyploid plant species on subsets of the genome indicate that unequal expression of the homologous alleles is quite common [18–20]. Therefore, the experimental capacity to discern the genomic origin of expressed homologous

* Corresponding authors.
E-mail addresses: xingwang.deng@yale.edu (X. Deng), ll4jn@virginia.edu (L. Li).
[1] These authors had equal contribution.

0888-7543/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2010.07.008

genes and quantify their regulated contribution to the transcript pool at the genome scale is much desired. Such capacity should facilitate various studies aimed at elucidating homologous gene regulation and the impact of polyploidy on genome evolution.

SNP is the most abundant form of DNA polymorphism and serves as a valuable molecular marker for genetic studies. Not surprisingly, much of the effort to experimentally distinguish the transcript of one allele from its highly similar counterpart has been directed toward SNPs in the coding regions. Experimental procedures based on several different principles to examine unequal transcription from SNP-defined alleles were successfully developed. One type of such methods employs the physical properties of the complementary DNA molecule such as mass [15] or DNA melting [21]. Another type of methods relies on enzymatic reactions that have different efficiency at the polymorphic sites. These include, for example, the RNase protection assay [22] and single nucleotide primer extension [12,23]. Microarray-based applications including mini-sequencing on microarray [24] and allele-specific microarray [14,25,26] were also developed.

Allele-specific microarrays typically involve pairs of probes with perfect match to one of the alleles. They provide a high throughput, multiplex, target-oriented platform for globally quantifying IAE that is affordable to most research laboratories. The power of this approach lies in the idea that a sequence mismatch between a probe and its target may significantly disrupt hybridization and attenuate that probe's signal. In developing genome scale allele-specific microarrays, a key technical consideration obviously is to design probes that offer sufficient sensitivity and specificity in discriminating the transcripts derived from different alleles. An equally critical yet less addressed consideration is that quantitative measurement of IAE requires the relative hybridization signal from the allele-specific probes to respond linearly to changes in the allele-specific transcript level. Because both the actual polymorphism and its sequence context will impact probe-target hybridization, a computation-based approach to select informative SNPs to discriminate expressed alleles remains insufficient. This may prove to be the major rate limiting step in the application of allele-specific microarrays in various species.

Advancement in high-density microarray technology permitted the development of tiling microarray that involves the representation of a genomic region with progressive oligonucleotide probes. Tiling arrays have been widely used in transcriptomic studies in plants [3,4,27–31]. Here we investigate the potential of tiling array in detecting and quantifying IAE focusing on 518 semi-randomly selected SNPs between two rice subspecies *indica* and *japonica*. Our effort revealed that over half of the SNPs are diagnostic and generate accurate IAE measurement in the reciprocal *japonica*×*indica* hybrids. Further analysis indicates that the tiling microarray-based experimental approach offers a versatile, target-oriented assay for examining IAE that can be readily applied in species for which DNA polymorphism information is available.

2. Results

2.1. Design of SNP-based tiling microarray

We chose rice to design allele-discriminating tiling microarray as both *indica* and *japonica* are completely sequenced and abundantly high quality SNPs identified [32–34]. In this study, we selected 518 SNPs between *indica* and *japonica* in the coding region of 475 genes (Table 1). To further verify the quality of these SNPs, we chose 21 that were predicted to result in the creation or disruption of a restriction site and performed cleaved amplified polymorphic sequence analysis. All 21 loci were verified in this analysis (Fig. S1), indicating that the annotated SNPs are a reliable source for designing probes specific to the *indica* or *japonica* alleles.

The schematic representation of the tiling strategy is illustrated in Fig. 1. The tiling design involves 25 sets (blocks) of 25-mer probes for each of the 518 SNPs. Within each block, which is used as a unit for measuring allelic output, there are two probes that match perfectly to the *indica* and *japonica* alleles, respectively. The other two probes, called mismatch probe, each contains one of the two remaining nucleotides at the SNP site. For each SNP, the 25 blocks differ in the position of the polymorphic sites, which were placed at all 25 possible positions within the probe (Fig. 1A). Thus all possible 25-mer probes spanning the polymorphic site for each SNP were included in the tiling design. The resultant 51,800 (518 times 100) probes were synthesized in triplicate (155,400 in total) in a single microarray, which was used throughout this study.

Table 1
Number of probe blocks, SNPs, and genes in the tiling microarray analysis.

	All	Diagnostic	IAE detected[a]	
			Japonica×*indica*	*Indica*×*japonica*
Blocks	12,590	519	115	85
SNPs	518	284	95	71
Genes	475	271	93	70

[a] In the F_1 hybrids. *japonica*×*indica*, *japonica* (♀)×*indica* (♂) F_1; and *indica*×*japonica*, *indica* (♀)×*japonica* (♂) F_1.

2.2. Identification of diagnostic blocks

To experimentally identify diagnostic probe blocks, we prepared a titration series consisting of five cDNA mixtures in which the *japonica* and *indica*-derived cDNA templates are mixed with the following proportions: 1:4, 1:2, 1:1, 2:1, and 4:1. We hybridized the SNP-based tiling microarray to this titration series and obtained five sets of hybridization data (Fig. 1B). For each block, we then calculated the relative abundance of the *indica*- and *japonica*-specific transcript (*i*/*j*) using the mismatch probes as background (see Methods). To score the 12,950 (528×25) blocks, we calculated the correlation coefficient (r) between the measured *i*/*j* ratio and the known *indica*/*japonica* ratio in the input cDNA across the titration series (Fig. 1C).

Individually measured *i*/*j* ratios do not necessarily reflect the input allelic ratio due to probe cross-hybridization and other confounding factors. Indeed, when all probes were considered, there was no obvious linearity between the measured and the input allele ratios and their values often vary to large extents (Fig. 2A). To screen for blocks that exhibit strong linearity at different input allelic ratio, we plotted the distribution of r values of the 12,950 blocks and observed a skewed normal distribution biased toward large positive values (Fig. 2B). This observation prompted us to perform the t test to examine whether an r value is significantly larger than the standard error, which led to the determination of an r value cutoff at 0.811 ($p = 0.05$; Fig. 2B). Using this cutoff, 519 of the 12,590 blocks were considered to be diagnostic for quantifying the linear effect of input cDNA on the measured allelic ratio (Fig. 2C). Further, the measured ratios from these blocks were also more comparable to the input ratio (comparing Fig. 2C with A). An example of a diagnostic block is illustrated where the measured ratios show an excellent correlation ($r > 0.99$) with the ratio of the template cDNA (Fig. 2D). Importantly, the 519 diagnostic blocks account for only 4.1% of all probe blocks tested yet they represent 284 (55%) of the 518 SNPs (Table 1). Together these results attest to the effectiveness of the tiling array method in uncovering diagnostic blocks for large numbers of SNPs.

2.3. Characterization of the diagnostic blocks

To gain insight into the property of the diagnostic blocks, we first investigated whether they are biased toward certain type of SNPs. In the 518 SNPs, transitions (A:G and C:T) are twice as frequent although there are twice as many possible transversions (A:C, A:T, C:G, and G:T; Fig. 3A). This pattern is highly similar to genome-wide SNP

Please cite this article as: H. He, et al., Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression, Genomics (2010), doi:10.1016/j.ygeno.2010.07.008

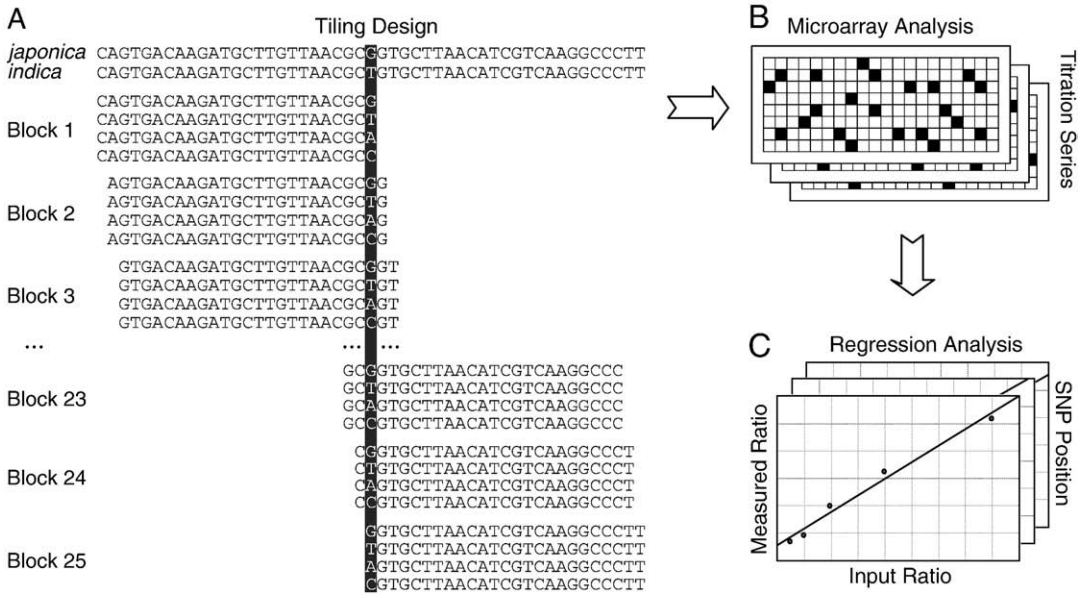

Fig. 1. Schematic representation of the tiling microarray approach to screen for diagnostic blocks for distinguishing allele-specific transcripts. (A) The tiling microarray involves 25 blocks of 25-mer oligonucleotide probes for each SNP. Within each block, four probes representing all four possible nucleotides at the SNP site are designed. Two of these probes match perfectly to the *indica* and the *japonica* alleles, while the other two are included as mismatch controls. (B) The tiling microarray is hybridized to a titration series consisting of five targets in which the *japonica* and *indica*-derived cDNA templates are mixed in various proportions. (C) The blocks are screened based on regression model for those that show strong linearity between the measured and the input ratio of allele-specific transcripts.

composition between the *japonica* and *indica* rice [32], consistent with the fact that the 518 SNPs were randomly chosen. When the nucleotide substitution pattern of the 284 SNPs was examined, we found that the frequency of the six types of changes is not significantly different from the 518 starting SNPs (chi square test, $p>0.1$; Fig. 3A). This analysis implies that SNP composition is not a major determinant for diagnostic blocks in the tiling array experiment.

We next examined the effect of SNP placement within the probe. Plotting the frequency of SNP position within the diagnostic blocks revealed a symmetric bimodal distribution, with the two modes locating at positions 6 and 20, respectively (Fig. 3B). This result indicates that the probability of identifying a diagnostic probe pair is a function of the SNP position within the probe. The center of the probe was typically the focus of genotyping and allele-specific microarrays

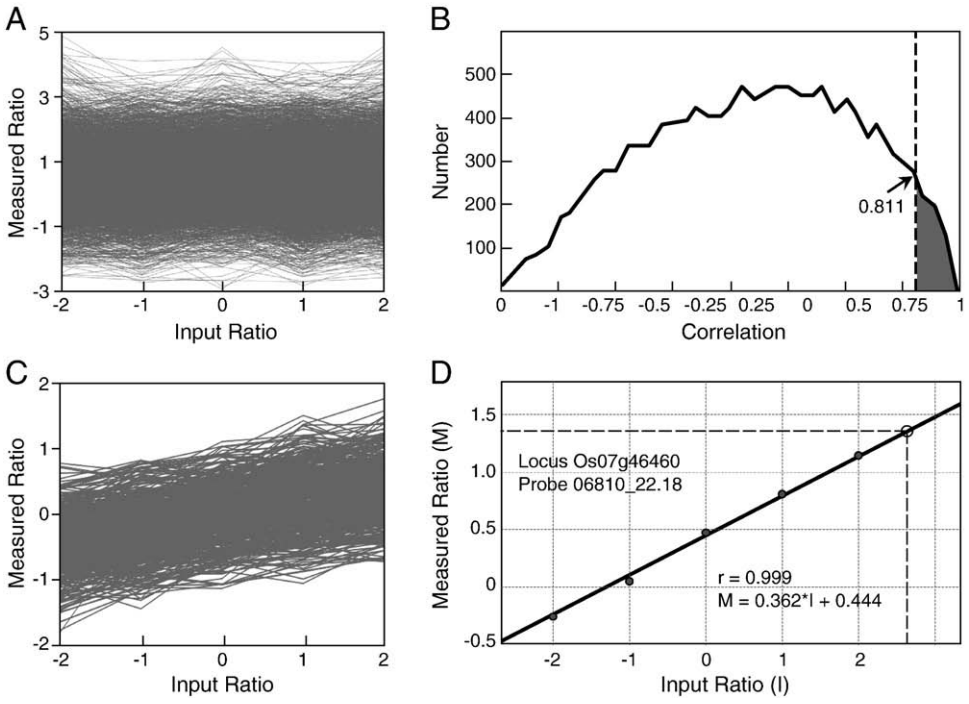

Fig. 2. Identification of diagnostic blocks for discriminating allele-specific transcripts. (A) Linear relation of the measured and input *indica*/*japonica* transcript ratio for all 12,900 blocks across the titration series. (B) A regression model is applied on the input and the measured *indica*/*japonica* transcript ratio and the correlation coefficient r for each block is calculated. The t test is performed to examine whether an r value is significantly larger than its standard error, which leads to an r value cutoff at 0.811 at $p=0.05$. (C) Linear relation of the measured and input *indica*/*japonica* transcript ratio for the 519 diagnostic blocks. (D) One example illustrating the linear relation between measured and input *indica*/*japonica* transcript ratio and demonstrating the use of the regression as a standard curve for deducing future *indica*/*japonica* transcript ratio at the given SNP.

Please cite this article as: H. He, et al., Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression, Genomics (2010), doi:10.1016/j.ygeno.2010.07.008

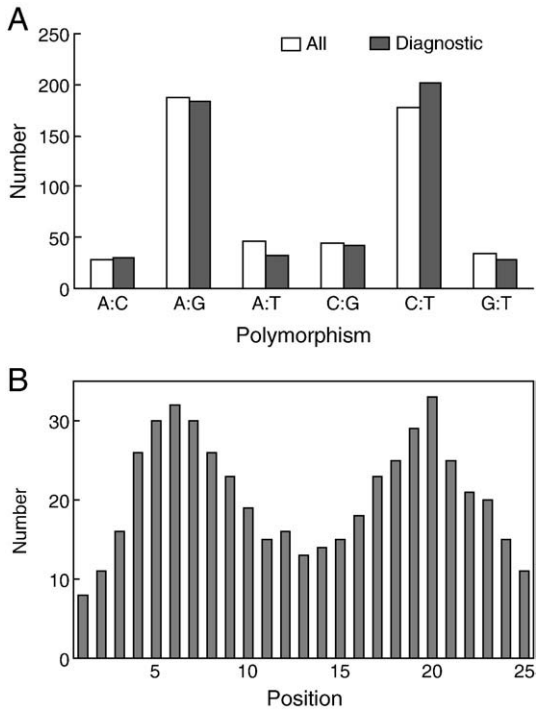

Fig. 3. Characterization of the diagnostic blocks. (A) Composition of all tested SNPs and those with at least one diagnostic block. (B) Position effect of SNP placement on the frequency of identifying diagnostic blocks. X axis is the position of SNP within the 25-mer probe. Y axis indicates the number of blocks that showed significant linear regression.

[20,26]. Surprisingly, we found that the center (position 13) is one of the least productive positions for diagnostic blocks (Fig. 3B), indicating that the use of fixed positions in microarray experiment will significantly limit the number of SNPs available for designing allele-discriminating probes. Together these results demonstrated the advantage of the tiling array strategy in interrogating large numbers of SNPs.

2.4. Determination of IAE in reciprocal rice hybrids

A corollary benefit of the tiling array method in that the regression used for selecting the diagnostic blocks doubles as a standard curve that can be used to deduce quantitatively the relative abundance of the allele-specific transcripts in real biological samples (Fig. 2D). As a proof-of-concept, we hybridized the tiling array to cDNA targets prepared from the leaf of F_1 plants derived from reciprocal crosses between *japonica* and *indica* rice. The measured allelic ratio at the transcript level was determined from the 519 diagnostic blocks and located at the individual standard curve. The relative abundance of the *japonica* and *indica* transcript in the hybrid plants was then deduced. The null hypothesis that both alleles of a given gene contribute equal amount of transcript was rejected (t test, FDR-adjusted $p<0.05$) for 115 and 85 blocks in the *japonica* (♀)×*indica* (♂) (Table S1) and *indica* (♀)×*japonica* (♂) (Table S2) hybrids, respectively. The 115 and 85 blocks represent 93 and 70 genes, respectively (Table 1). Out of 271 genes interrogated by at least one diagnostic probe block, 93 (34%) and 70 (26%) exhibited significant deviation from equal expression of the two alleles in one of the reciprocal F_1 hybrids (Table 1). Thus, approximately one third of the examined rice genes exhibit IAE in the leaf of hybrid rice plants.

There are a number of cases where a gene is interrogated by more than one diagnostic block. For example, in the *japonica* (♀)×*indica* (♂) hybrid, 17 genes are covered by at least two diagnostic blocks (Table S1). In the vast majority of cases, the different blocks produced consistent results. Only in the case of Os10g33710 did the two blocks interrogating the same SNP generate inconsistent IAE measurements. It should be noted that each of the block produced consistent result in the reciprocal crosses (Tables S1 and S2), suggesting that the discrepancy at this SNP was likely caused by cross-hybridization of one of the two blocks rather than experimental fluctuation. We further compared the genes that exhibit IAE in the reciprocal rice hybrids and found significant overlapping between the two gene sets (Fig. 4). Closer inspection revealed that the vast majority (93%) of the overlapping genes exhibit the same direction of imbalanced expression in both hybrids (Fig. 4). Together, our results indicate that the tiling array method is effective at interrogating large numbers of SNPs in real biological samples and producing useful information on IAE of the corresponding genes.

2.5. Experimental validation of tiling array-detected IAE

We carried out two sets of independent experiment to validate the tiling array-detected IAE in the rice hybrids. First, we quantified allele level gene expression by performing an allele-discriminating real-time PCR assay on reverse transcribed cDNA in the *japonica* (♀)×*indica* (♂) hybrid. In this assay, SNP-bearing cDNAs are distinguished with a pair of allele-specific forward primers and a common reverse primer. Using these primers, abundance of allele-specific transcripts is correlated with the yield of the respective PCR amplicon [35]. We selected five SNPs for validation, including four with IAE and one with allele-neutral expression according to the tiling array analysis. We found that the results from the PCR assay are in strong agreement with the tiling array measurements (Fig. 5). For example, at the SNP063644 site (Os02g33560), we estimated from the tiling array experiment that the *japonica* allele is present at a 0.32 to 1 ratio against the *indica* allele in the F_1 hybrid. This ratio is very close to that deduced in the PCR analysis (0.38; Fig. 5). As another example, both the tiling array (0.96) and PCR (0.91) experiments detected allele-neutral expression at the SNP102045 (Os07g37100) site (Fig. 5).

Previously, He et al. [36] used RNA-sequencing (RNA-Seq) to examine allele-specific transcripts from 20,638 homologous genes between *japonica* and *indica* by comparing the sequencing reads at each SNP-bearing exons. We compared the results from tiling array with the IAE profile determined by the RNA-Seq data. Of the 95 SNPs with detected IAE by tiling array, we were able to find at least one allele-specific sequence read for 20 SNPs. Based on the p-value of the IAE in the RNA-Seq data, the 20 SNPs were divided into three groups (Table 2). Group I consists of 10 SNPs with significant and consistent

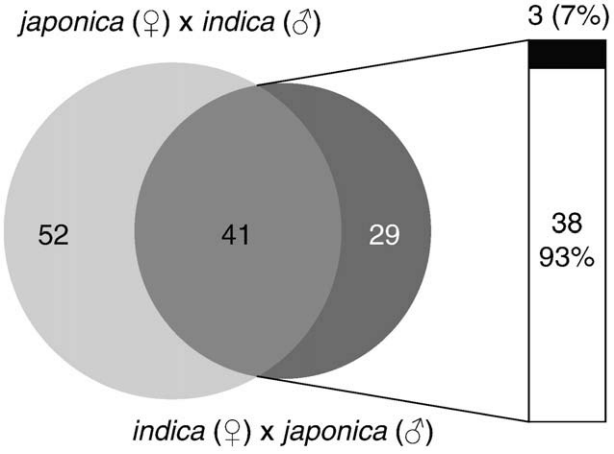

Fig. 4. IAE in reciprocal rice F_1 hybrids. Genes exhibiting IAE in the reciprocal hybrids are compared by Venn diagram. Among genes with detected IAE in both hybrids, the direction of transcript bias is also analyzed.

Please cite this article as: H. He, et al., Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression, Genomics (2010), doi:10.1016/j.ygeno.2010.07.008

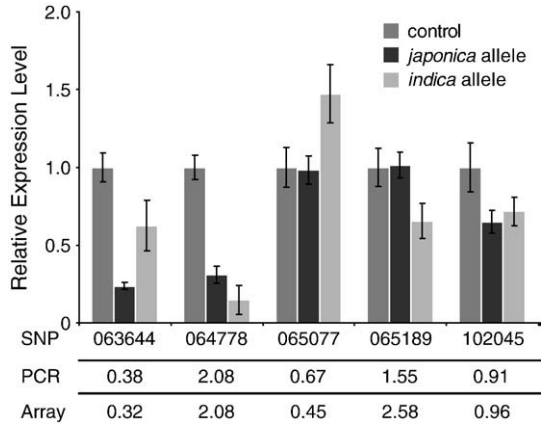

Fig. 5. Validation of IAE by allele-discriminating real-time PCR. Five SNPs, SNP063644 (Os02g33560), SNP064778 (Os10g32880), SNP065077 (Os02g58340), SNP065189 (Os06g01360), and SNP102045 (Os07g37100) are selected. Primer sets are designed to distinguish alleles in the *japonica* (♀)×*indica* (♂) F_1 hybrid and used in real-time reverse transcription coupled PCR. The ratio of allele-specific PCR products is calculated after normalization against the control amplicon and compared with the tiling array-deduced *japonica*/*indica* allelic ratio.

IAE detected by both methods. In group II, both methods detected the same directional bias of expression at the SNPs though RNA-Seq data lacks statistic significance. For the two SNPs in group III, the two methods reported inconsistent results, which apparently can be attributed to the sparse sequence reads spanning these SNPs (Table 2). Together these results indicate that tiling array-based high throughput detection of IAE is quantitative and reliable.

2.6. IAE and differential gene expression in hybrid

Our final objective is to demonstrate the utility of IAE in the context of differential gene expression in the hybrids, which has been implicated in the phenotypic manifestation of heterosis [37–39]. Comparing with results from a previous microarray study of the *japonica* (♀)×*indica* (♂) cross in which differential expression of 2416 (7%) genes was detected in the F_1 based on significant pair wise comparisons [39], we found that 17 (18%) genes showing IAE are differentially expressed (Fig. 6A). Transcripts in the hybrid could accumulate to the mid-parent level (additivity), the high or low-parent level (high- or low-parent dominance), or the level above the high-parent (over-dominance) or below the low-parent (under-dominance) [38]. Interestingly, while 29% of the genes exhibited a pattern that could be distinguished from additivity at the genome level [39], all 17 genes with IAE exhibited the non-additive gene action (Fig. 6A). Further, alignment of the promoter regions of these 17 genes from both *indica* and *japonica* identified *cis*-variations for 11 genes based on a set of known transcription factor binding sites [40].

Combining IAE, gene level differential expression and variation in *cis*-regulatory sequences allowed us to gain mechanistic insight into gene regulation in the hybrids. A specific example is illustrated in Fig. 6b. The Os01g31110 gene exhibits under-dominance pattern with an expression level in the hybrid approximately 70% relative to both parental lines (Fig. 6B). In the hybrid, the *indica* allele contributes much less to the transcript pool and averaging the two alleles recapitulates the under-dominance expression pattern (Fig. 6B). These results indicate that down regulation of this gene in the hybrid is caused by a specific repression of the *indica* allele. Examination of the highly homologous promoter regions between the two parental lines revealed that *indica* harbors an additional putative *cis*-element resembling the GCC box (Fig. 6C). It is thus possible that the cryptic *cis*-element is recognized in the hybrid, which contains a different set of *trans*-factors than *indica*, to confer repression. Together our results demonstrate that IAE is an excellent read out for the complex interactions between *cis*-regulatory elements and *trans*-acting regulators that lead to non-additive gene expression in the hybrid.

Table 2
Comparison of IAE measured by tiling microarray and RNA-sequencing.

	Gene	SNP	*Japonica* allele	*Indica* allele	Measured ratio (*j*/*i*)[a]	Reads (*j*:*i*)[b]	p-value[c]
Group I	Os02g25900	066598_1	A	G	0.05	0:16	0.00002
	Os01g42980	073157_1	C	T	6.03	23:5	0.00046
	Os01g25600	072081_2	T	C	0.20	6:22	0.00186
	Os03g17520	070093_3	A	G	0.36	15:35	0.00330
	Os08g33830	067548_4	C	A	2.68	28:11	0.00474
	Os01g02020	103194_1	A	G	0.23	4:16	0.00591
	Os03g47930	065067_3	T	G	5.71	14:3	0.00636
	Os02g54030	058942_1	G	A	0.39	12:28	0.00829
	Os03g58780	105727_1	G	A	9.27	8:1	0.01953
	Os01g67770	062250_1	G	C	0.32	1:7	0.03516
Group II	Os03g09280	073386_1	C	T	0.21	0:4	0.06250
	Os01g63820	072005_1	T	C	2.48	31:21	0.10580
	Os05g34720	067880_2	A	G	0.15	1:5	0.10938
	Os01g70570	068196_3	C	T	0.28	0:3	0.12500
	Os09g33670	060504_1	A	T	3.90	14:9	0.20244
	Os03g15870	061468_4	C	T	0.29	5:9	0.21198
	Os02g35190	066353_2	T	C	3.37	6:3	0.25391
	Os04g35290	069367_2	A	G	0.38	7:9	0.40181
Group III	Os03g60920	059711_1	G	A	3.80	0:1	0.50000
	Os06g49010	073179_1	C	T	0.48	1:1	0.75000

[a] SNPs with significant IAE determined by tiling array analysis.
[b] Number of allele-specific sequence reads spanning a given SNP as obtained from He et al. [36].
[c] Binomial test with the null hypothesis that two parental alleles are expressed equally as described in He et al. [36].

3. Discussion

DNA microarrays with SNP-discriminating probes provide an effective platform for the simultaneous measurement of allele-specific expression for large numbers of genes [14,20,25,26]. We deem that there are two technical considerations critical to the broad application of this method in diverse experimental systems. One is the design of probes that offer sufficient sensitivity in discriminating the transcripts derived from different alleles. The other is the need to guarantee that the relative hybridization signal from the allele-specific probes is linearly proportional to the actual allele-specific transcript level. In the current work, we used a strategy coupling a tiling design with hybridization to a premixed series of *japonica*/*indica* cDNA with known proportions to address both considerations (Fig. 1). We then developed a bioinformatic method based on a linear regression model to experimentally screen for the most diagnostic probe blocks that interrogate a given SNP (Fig. 2).

Our results help to overcome two major limitations in the current SNP-based allele-specific microarray, which is based on the idea that a sequence mismatch between a probe and its target significantly attenuates its signal and has been heavily influenced by genotyping techniques. First, it is generally assumed that the allele-specific probes would hybridize better to the corresponding transcript than the transcript with a mismatch. Thus, the focus of most previous studies was on probes that show significant deviation from comparable hybridization signals [14,20,25,26]. However, there was typically no internal control for whether the deviation is caused by other compounding factors such as cross-hybridization. The test for probes with unequal signals only provides information on the directional change but not the degree of IAE and lacks the statistic power for boarder line genes. These could be reasons for ambiguous IAE calling for many loci, which could not be validated with low throughput experimental techniques [26]. In contrast, our method relies on hybridization to a titration series of known allelic input and a linear regression model to experimentally screen for the most diagnostic probe sets. Further, the regression doubles as a standard curve for

Please cite this article as: H. He, et al., Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression, Genomics (2010), doi:10.1016/j.ygeno.2010.07.008

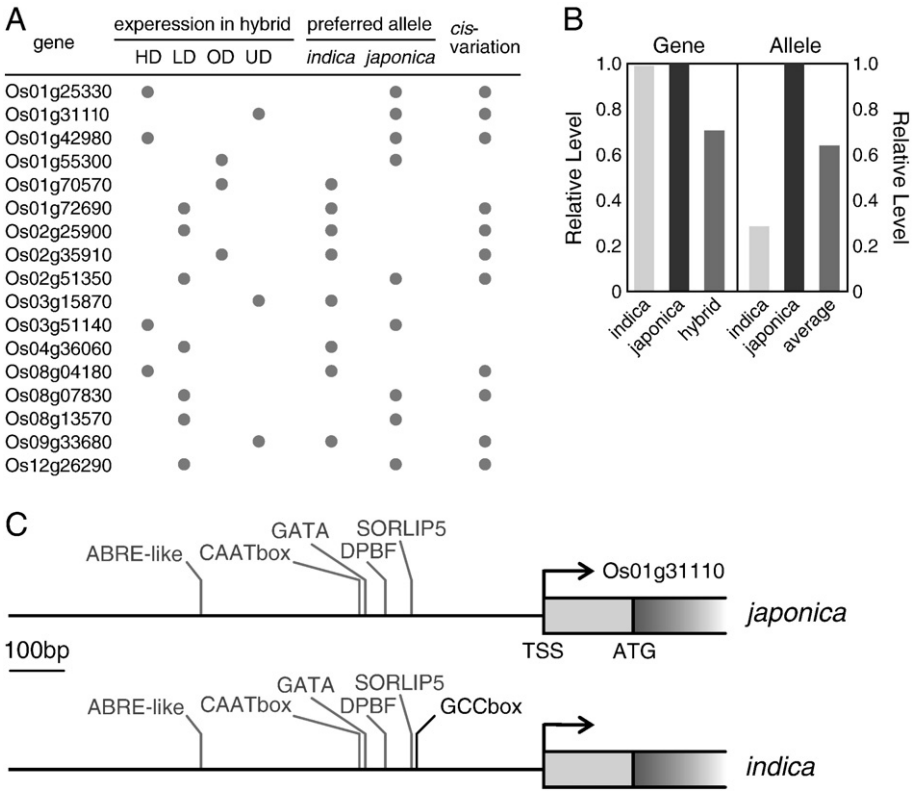

Fig. 6. Integrating IAE and *cis*-variation in differential gene expression analysis. (A) Analysis of IAE, differential expression, and *cis*-variation in the *japonica* (♀)×*indica* (♂) cross. Expression patterns were obtained from Zhang et al. [39]. HD, high-parent dominance. LD, low-parent dominance. OD, over-dominance. UD, under-dominance. (B) Shown on the left is Os01g31110 expression at the gene level in *indica*, *japonica* and their hybrid after normalization against the *japonica* level. Relative expression in the hybrid from the *indica* and *japonica* alleles and the average is shown on the right. IAE measurement is the mean of three probe sets interrogating the same SNP that can be found in Table S1. (C) Comparison of the putative *cis*-elements in the promoter region of Os01g31110 between *japonica* and *indica*. The 1 kb upstream region from the transcription start site (TSS) was used as the promoter.

quantitatively deducing the relative abundance of allele-specific transcripts in future experiments (Fig. 2), thus providing both a reliable and quantitative means for measuring allele-specific gene expression.

Second, in most allele-specific arrays the polymorphic site is placed at a fixed position, typically at the center of the probe, with the goal to maximize the allele discrimination power of the probes [20,26]. In several previous studies, the impact of the position of the polymorphic site within the probe on the ability to distinguish alleles at the RNA level was examined [14,25]. At a 1:1 allelic input ratio, the probability of detecting a single base change was shown to be a function of position within the probe [25]. Consistent with this conclusion, we found that the effect of SNP placement on the frequency of identifying diagnostic blocks is a symmetric bimodal distribution with the two modes locating at positions 6 and 20 within the 25-mer probe, respectively (Fig. 3B). This result indicates that the probability of SNP detection at the transcript level is a function of the position of SNP and highlights the unique advantage of the tiling microarray strategy in identifying probes with broad coverage of allele-discriminating SNPs in the genome.

The tiling array method proves to be an effective approach for selecting informative SNPs in quantifying IAE. We successfully identified a total of 519 diagnostic blocks representing 284 (55%) of the 518 tested SNPs (Table 1). Characterization of these 284 SNPs revealed no obvious compositional bias (Fig. 3A). We note that in seedling leaves, the only organ used in this study, only three quarters of the genes may be expressed [39,41]. Thus, a portion of the SNPs failed to produce diagnostic blocks which could be due to a lack of detectable transcript. Taken together, our results indicate that tiling array method is able to identify diagnostic blocks for a majority of *indica–japonica* SNPs and hence has the potential to provide coverage for the whole genome. Coupled with the ability to design multiplex assays on high-density oligonucleotide array, the tiling method is ideal to analyze allele-specific gene expression in multiple samples with sufficient biological replicates in a truly high throughput manner.

With the 519 diagnostic blocks, we analyzed IAE in the leaf of rice F_1 hybrids. Out of the 271 genes interrogated by at least one diagnostic block, we found that 93 (34%) and 70 (26%) genes exhibit significant deviation from equal expression of the two parental alleles in the *japonica* (♀)×*indica* (♂) and *indica* (♀)×*japonica* (♂) hybrids, respectively (Fig. 4). These results are strongly supported both by available RNA-Seq data (Table 2) and allele-specific PCR analysis (Fig. 5), demonstrating the reliability of the tiling array method. The proportions of rice genes exhibiting IAE are consistent with the frequency of IAE detected in previous studies in maize hybrids focusing on different sets of genes [10,15,16]. This observation implies that wide occurring of IAE is common in heterozygous plants. Our results also showed that for some genes, IAE is so strong that one parental allele is predominantly expressed. This could be a mechanism associated with non-additive gene actions in the hybrids (Fig. 6A). Thus, global detection of IAE is of considerable interest in characterizing transcriptional control in plant genomes by helping to identify changes in *cis*-regulatory sequences or chromatin states between different genotypes.

RNA-Seq using next-generation deep sequencing technologies can be used to directly measure allele-specific transcripts [36]. In principle, RNA-Seq not only yields information on the existence of heterozygous transcripts, but also help estimate the expression level of each allele through quantifying and normalizing allele-specific tags. However, the effectiveness of this method in global quantification of IAE has not been systemically tested. In a recent work, it was shown that current tag

Please cite this article as: H. He, et al., Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression, Genomics (2010), doi:10.1016/j.ygeno.2010.07.008

density (millions to tens of millions of tags per transcriptome) is insufficient to deduce quantitative differences among most known RNA isoforms due in large parts to the sparseness of tags specific to splice junctions [5]. It is therefore possible that the tag density in current RNA-Seq studies will fall short of providing quantitative information for many heterozygous transcripts. This appears to be the case in our comparison of IAE measurement by tiling array analysis and available RNA-Seq data from comparable samples (Table 2). In contrast to the open ended nature of RNA-Seq, tiling array represents an economic alternative platform for target-oriented measurement of IAE at the genome scale in diverse experimental systems.

4. Methods

4.1. Plant materials

Rice strains 93–11 (*Oryza sativa* L. ssp *indica* cultivar), Nipponbare (*O. sativa* L. ssp. *japonica* cultivar), and their reciprocal F_1 hybrids were used in this study. The F1 hybrids were generated and characterized as previously described [39]. The shoot tissue from seedling grown in environmentally controlled growth chambers at 28 °C were harvested at the four-leaf stage, frozen in liquid nitrogen and homogenized. Total RNA and polyA (+) RNA were subsequently isolated using the RNeasy Plant Mini kit (Qiagen) and the Oligotex mRNA kit (Qiagen) according to the manufacturer's recommendations, respectively.

4.2. SNP selection

Initial SNP information was downloaded from the BGI Rice Information System (http://rice.genomics.org.cn/rice/link/download.jsp). A set of approximately 3000 SNPs were randomly selected and confirmed by a BLAT search on *japonica* and *indica* genome sequences. The 518 SNPs used in this study were then selected by a series of criteria: (a) the candidate SNPs should locate in coding regions supported by rice full-length cDNA clones with a size range between 1 and 3 kp; (b) the polymorphic nucleotides should have a sequencing quality score no less than 90 according to Yu et al. [34]; (c) two adjacent SNP sites should be at least 100 bp from each other; (d) the 25-mer probes covering the SNP site should have a GC content range of 40% to 60%, and melting temperature in the range of 70 ± 5 °C; (e) similarity search was performed by a BLAST search of the candidate probes against the *japonica* rice cDNA annotation (http://rice.plantbiology.msu.edu/data_download.shtml) to avoid sequences with multiple copies; and (f) synonymous and non-synonymous SNPs were analyzed so they are represented at comparable frequency.

4.3. Tiling microarray design and experiment

For each selected SNP, 25 blocks of 25-mer oligonucleotide probes were designed. Within each block, four probes representing all four possible nucleotides at the SNP site were included. Two of these probes match perfectly to the *indica* and the *japonica* alleles, while the other two are mismatch controls. The 25 blocks differ in the position of the polymorphic site that was placed at all 25 possible positions. Thus, 100 probes spanning the polymorphic site were included for each SNP. The resultant 51,800 probes were synthesized in triplicate in a single microarray produced on the Maskless Array Synthesizer platform as previously described [4]. Poly(A+) RNA from *japonica*, *indica* and their reciprocal hybrids was reverse transcribed using an oligo$(dT)_{18}$ primer, during which amino-allyl-modified dUTP (aa-dUTP) was incorporated. The aa-dUTP decorated cDNA was fluorescent labeled by conjugating the monofunctional Cy3 dye (GE Healthcare) to the amino-allyl functional groups in the cDNA. Dye-labeled target was quantified using a spectrophotometer and 2 μg used for hybridization to each array. In the titration series experiment, the dye-labeled targets from *japonica* and *indica* were mixed in known proportions before hybridization. Microarray design and experimental data are available in the NCBI Gene Expression Omnibus (GSE20678).

4.4. Computational analysis of microarray data

In the first microarray experiment involving the titration series, the obtained hybridization intensity was log2-transformed and quantile-normalized across the five arrays. For each block, the average of the two mismatch probes was used as the background value. The intensity for the allele-specific probe (i or j) was computed by subtracting the background value from the observed intensities. A baseline value of 0 was used when negative values were encountered. The array-measured relative abundance of the *indica*- and *japonica*-specific transcript (M) was denoted by the i/j ratio. A regression method was then applied on the input ratio (I) and the measured ratio M: $M = a + b\ I$, where a means intercept on the Y axis and b means the slope. To score the 12,950 (528×25) blocks, we calculated the correlation coefficient (r) between M and I for each block across the titration series. We performed the t test to examine whether an r value is significantly larger than its standard error and obtained an FDR-adjusted p-value for each block. Finally, three criteria were used to select the diagnostic blocks for which (i) FDR p-value<0.05; (ii) $-1<a<1$; and (iii) $b>0.1$.

In the second microarray experiment involving the reciprocal rice hybrids, the obtained hybridization intensity was log2-transformed and quantile-normalized across the three replicates. The measured ratio M was determined from the diagnostic blocks in the reciprocal hybrids as described earlier. The input ratio I was computed from M based on the established parameters (a, b) of the regression equations. Log2-transformed I values from the three replicates for a given block were used to perform the one-sample t test to examine whether the mean score differs from 0, with the null hypothesis that there is no significant difference. After obtaining p-values, we determined their rank in ascending order: Rank(i) where $i = 1$ to n. The p-values were then adjusted following the formula $FDR(i) = \min(Pvalue(i)\ n/Rank(i), FDR(i+1))$. This process was reiterated as i was moved from n to 1 until $FDR(n) = Pvalue(n)$. The blocks with FDR-adjusted p-value less than 0.05 were selected as having detected imbalanced allelic expression.

4.5. Analysis of cis-variation

The full-length cDNA from the 17 *japonica* genes shown in Fig. 6A were mapped to the *indica* genome assembly by BLAT. Putative orthologous genes were identified when the mapped length was ≥95%. The start codon of the annotated *japonica* gene was used as the anchoring point to extract the upstream 1.5 kb sequences from both the *japonica* and *indica* genome. These sequences or the upstream 1 kb sequences from the transcription start site were used as putative promoter regions. The promoter sequences were aligned by BLASTN to validate homology ($E \leq 1.0e^{-5}$ and identify ≥95%). Homologous promoter regions were then scanned for putative transcription factor binding motifs using the position weight matrices of 99 known binding motifs constructed in *Arabidopsis* [40]. The log-likelihood scoring function and associated threshold scores for individual matrices were the same as described by Megraw et al. [40].

4.6. Real-time allele-discriminating PCR

We adopted the simple allele-discriminating PCR method [35] to examine the abundance of allele-specific transcripts in the F_1 hybrids. For each pair of SNP-defined alleles, three primers were used that include two allele-discriminating forward primers and a common reverse primer. The two allele-discriminating forward primers are designed so that the last base at the 3′ end corresponds to the specific nucleotide for each allele. Importantly, each primer incorporates one additional mismatch at the penultimate base, resulting in one mismatch

Please cite this article as: H. He, et al., Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression, Genomics (2010), doi:10.1016/j.ygeno.2010.07.008

between the primer and its target template but two mismatches for the non-target template [35]. In addition, a common forward primer is used with the common reverse primer as a control for overall transcript abundance. Primer sequences are listed in Table S3.

For PCR, first-strand cDNA was synthesized from 4 μg of total RNA in a volume of 20 μl, using the SuperScript II first-strand synthesis system and an Oligo$(dT)_{18}$ primer (Invitrogen). Quantitative PCR was performed on an ABI 7500 real-time PCR system using a SYBR Green kit (Applied Biosystems). Conditions for the reaction were as follows: 1 cycle at 95 °C for 10 min and 40 cycles at 95 °C for 15 s, 58 °C for 20 s and 72 °C for 30 s. Following PCR, samples were subjected to a melting analysis to confirm specificity of the amplicon. PCR for each primer set was done at least three times and the mean reported. Relative abundance of allele-derived transcripts was determined by normalizing against the control amplicon that targets the identical region between the two parental alleles. The ratio between the *japonica* and *indica* alleles in the transcript pool was then calculated based on the relative abundance.

Supplementary data to this article can be found online at doi:10.1016/j.ygeno.2010.07.008.

Acknowledgments

We thank Farah Steele for her technical assistance. This work was supported in part by a NSF Plant Genome Program grant (DBI-0922604) to XWD, a NSF Plant Genome Program grant (DBI-0922526) to LL, and grants from the Ministry of Science and Technology of China (2009DFB30030) and the Ministry of Agriculture of China (2009ZX08012-021B).

References

[1] D.J. Lockhart, E.A. Winzeler, Genomics, gene expression and DNA arrays, Nature 405 (2000) 827–836.
[2] B.C. Meyers, D.W. Galbraith, T. Nelson, V. Agrawal, Methods for transcriptional profiling in plants. Be fruitful and replicate, Plant Physiol. 135 (2004) 637–652.
[3] K. Yamada, et al., Empirical analysis of transcriptional activity in the *Arabidopsis* genome, Science 302 (2003) 842–846.
[4] L. Li, et al., Genome-wide transcription analyses in rice using tiling microarrays, Nat. Genet. 38 (2006) 124–129.
[5] H. Li, et al., Determination of tag density required for digital transcriptome analysis: application to an androgen-sensitive prostate cancer model, Proc. Natl Acad. Sci. USA 105 (2008) 20179–20184.
[6] S.M. Brady, et al., A high-resolution root spatiotemporal map reveals dominant expression patterns, Science 318 (2007) 801–806.
[7] Y.L. Jiao, et al., A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies, Nat. Genet. 41 (2009) 258–263.
[8] P.R. Buckland, Allele-specific gene expression in humans, Hum. Mol. Genet. 13 (2004) 255–260.
[9] L. Milani, et al., Allelic imbalance in gene expression as a guide to *cis*-acting regulatory single nucleotide polymorphisms in cancer cells, Nucleic Acids Res. 35 (2007) e34.
[10] N.M. Springer, R.M. Stupar, Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize, Plant Cell 19 (2007) 2391–2402.
[11] H. Khatib, Is it genomic imprinting or preferential expression? Bioessays 29 (2007) 1022–1028.
[12] H. Yan, W. Yuan, V.E. Velculescu, B. Vogelstein, K.W. Kinzler, Allelic variation in human gene expression, Science 297 (2002) 1143.
[13] N.J. Bray, P.R. Buckland, M.J. Owen, M.C. O'Donovan, *Cis*-acting variation in the expression of a high proportion of genes in human brain, Hum. Genet. 113 (2003) 149–153.
[14] H.S. Lo, et al., Allelic variation in gene expression is common in the human genome, Genome Res. 13 (2003) 1855–1862.
[15] M. Guo, et al., Allelic variation of gene expression in maize hybrids, Plant Cell 16 (2004) 1707–1716.
[16] R.M. Stupar, N.M. Springer, *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid, Genetics 173 (2006) 2199–2210.
[17] I.J. Leitch, M.D. Bennett, Polyploidy in angiosperms, Trends Plant Sci. 2 (1997) 470–476.
[18] K.L. Adams, R. Cronn, R. Percifield, J.F. Wendel, Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing, Proc. Natl Acad. Sci. USA 100 (2003) 4649–4654.
[19] J. Wang, et al., Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids, Genetics 172 (2006) 507–517.
[20] L. Flagel, J.A. Udall, D. Nettleton, J.F. Wendel, Duplicate gene expression in allopolyploid Gossypium reveals two temporally distinct phases of expression evolution, BMC Biol. 6 (2008) 16.
[21] S. Jeong, Y. Hahn, Q. Rong, K. Pfeifer, Accurate quantitation of allele-specific expression patterns by analysis of DNA melting, Genome Res. 17 (2007) 1093–1100.
[22] E. Winter, F. Yamamoto, C. Almoguera, M. Perucho, A method to detect and characterize point mutations in transcribed genes: amplification and overexpression of the mutant c-Ki-ras allele in human tumor cells, Proc. Natl Acad. Sci. USA 82 (1985) 7575–7579.
[23] J. Singer-Sam, V. Chapman, J.M. LeBon, A.D. Riggs, Parental imprinting studied by allele-specific primer extension after PCR: paternal X chromosome-linked genes are transcribed prior to preferential paternal X chromosome inactivation, Proc. Natl Acad. Sci. USA 89 (1992) 10469–10473.
[24] U. Liljedahl, M. Fredriksson, A. Dahlgren, A.C. Syvänen, Detecting imbalanced expression of SNP alleles by minisequencing on microarrays, BMC Biotechnol. 4 (2004) 24.
[25] J. Ronald, J.M. Akey, J. Whittle, E.N. Smith, G. Yvert, L. Kruglyak, Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays, Genome Res. 15 (2005) 284–291.
[26] J.A. Udall, J.M. Swanson, D. Nettleton, R.J. Percifield, J.F. Wendel, A novel approach for characterizing expression levels of genes duplicated by polyploidy, Genetics 173 (2006) 1823–1827.
[27] L. Li, et al., Transcriptional analysis of highly syntenic regions between *Medicago truncatula* and *Glycine max* using tiling microarrays, Genome Biol. 9 (2008) R57.
[28] L. Li, et al., Global identification and characterization of transcriptionally active regions in the rice genome, PLoS ONE 2 (2007) e294.
[29] L. Li, et al., Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture, Genome Biol. 6 (2005) R52.
[30] S.P. Hazen, et al., Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays, Genome Biol. 10 (2009) R17.
[31] Y. Kurihara, et al., Transcriptome analyses revealed diverse expression changes in ago1 and hyl1 *Arabidopsis* mutants, Plant Cell Physiol. 50 (2009) 1715–1720.
[32] F.A. Feltus, et al., An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments, Genome Res. 14 (2004) 1812–1819.
[33] Y.J. Shen, et al., Development of genome-wide DNA polymorphism database for map-based cloning of rice genes, Plant Physiol. 135 (2004) 1198–1205.
[34] J. Yu, et al., The genomes of *Oryza sativa*: a history of duplications, PLoS Biol. 3 (2005) 38.
[35] M. Bui, Z.C. Liu, Simple allele-discriminating PCR for cost-effective and rapid genotyping and mapping, Plant Meth. 5 (2009) 1.
[36] G.M. He, et al., Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids, Plant Cell 22 (2010) 17–33.
[37] S. Song, H. Qu, C. Chen, S. Hu, J. Yu, Differential gene expression in an elite hybrid rice cultivar (*Oryza sativa*, L) and its parental lines based on SAGE data, BMC Plant Biol. 7 (2007) 49.
[38] R.A. Swanson-Wagner, et al., All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents, Proc. Natl Acad. Sci. USA 103 (2006) 6805–6810.
[39] H.Y. Zhang, et al., Genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids, Mol. Plant 1 (2008) 720–731.
[40] M. Megraw, et al., MicroRNA promoter element discovery in Arabidopsis, RNA 12 (2006) 1612–1619.
[41] M. Schmid, et al., A gene expression map of *Arabidopsis thaliana* development, Nat. Genet. 37 (2005) 501–506.

Please cite this article as: H. He, et al., Development of a versatile, target-oriented tiling microarray assay for measuring allele-specific gene expression, Genomics (2010), doi:10.1016/j.ygeno.2010.07.008